



**Statistical thinking  
foundations for  
postgraduates across  
disciplines**

***Helen MacGillivray***

**QUT**



# Outline of presentation

- **QUT Maths Access Centre (QUTMAC) & how it got involved in support for pg's**
- **Trial & evaluation, 2005**
- **Symposia in Statistical Thinking for Postgraduates**
- **Associated developments, 2006-current**
- **Some pre-symposia questionnaire results**
- **Some participant comments**
- **Some observations**

# **QUT Maths Access Centre established 2004**

**Aim: to develop and provide university-wide learning support for students in any area for which numeracy, mathematical and/or statistical confidence are needed.**

**The QUTMAC programmes now include:**

- **drop-in facilities (3 campuses), with resources & schedule of duty tutors.**
- **student-driven support sessions for a range of courses or topics.**
- **provision & development of general and course-specific resources and questionnaires for student self-diagnostic support.**
- **current areas supported include engineering, chemistry, nursing, education, human movement, psychology, maths, stats**
- **an apprenticeship-model program with mentoring and training for talented undergraduates to develop tutoring and communication skills in mathematics and statistics.**
- **collaboration with academic & learning support programs.**
- **a website (Blackboard) and a Facebook.**
- **data collection and analysis.**

**QUTMAC clearly oriented to helping at introductory level: first years & anyone with lack of confidence in background or foundation skills**

**But**

**In agreeing to fund it (very modestly @ \$80k pa; current cost much more & we are in debt), DVC recommended “doing something” for postgraduates**

**!!**

**In 2005, Manager of Research Students Centre approached me asking if I could help in statistics for commencing postgraduates.**

**Me, very warily:**

- \* *pg's tend to say they want individual help, &***
  - \* *people often think pg's need workshops in specific advanced methods,***
- but, before any of that, my experience is that***
- \* *they need to understand role of statistics in research, &***
  - \* *they need sufficient understanding of core statistical data analysis methods to have some confidence for their research & possibly discipline-specific or more advanced methods***

**Manager: *exactly!***

**(*me, thinks: wonderful to have someone who understands*)**

**So me & the manager plotted & planned a trial to see if  
“*Statistical Thinking for Postgraduates Across Disciplines*”  
would fly.**

# Trial & evaluation, 2005

- Offered trial initial session, described by
- ***Session Objectives: Learning about***
  - *Your statistical thinking*
  - *Planning a statistics-friendly data investigation.*
  - *Statistical questions and research questions.*
  - *Choosing, using and interpreting statistical graphs and procedures*
- ***The session will refer to real data investigations in contexts that do not depend on discipline-specific knowledge.***
- **A pre-session “diagnostic” questionnaire sent to registrants to help them (& me!)**
  - included stats literacy questions from Therese Wilson’s PhD (for 1<sup>st</sup> years) & questions on core stats methods oriented to misconceptions & even mistakes often seen in pg/research in other disciplines
  - see later for details & some interesting results

# Trial & evaluation, 2005

- **74 registered, from all faculties, 51 attended**
  - Included staff, particularly from Health
- **Focus on**
  - Planning a study
  - Importance of identifying variables & their types, and subjects (“design your spreadsheet for your raw data”)
  - Turning research questions into statistical questions
  - General principles of estimation & error of estimation, prediction, modelling, fitting, testing, checking assumptions
  - Choice of graphs
  - What is needed to choose analysis method(s)

# Trial & evaluation, 2005

- Evaluation included asking what they'd like for next session & in future
- 38 evaluations (across faculties) with lots of comments
- 66% agreed or strongly agreed it addressed their information needs
  - Surprisingly large given it was just an introduction
- 85-90% agreed or strongly agreed that
  - Objectives were clear
  - Content was relevant
  - Well structured & presented
  - They learned something useful to their pg research



# Trial & evaluation, 2005

- **Comments generally indicated that they**
  - wanted more
  - ranged from very little background (“What’s a t-test?” “What’s ANOVA?”) to having one ug stats course
  - wanted basics: “chisq, ANOVA, regression, confounders,” “what to use when & where” “when to use different graphs”
- **General comments such as:**
  - “a pleasure to attend, very informative & delivered in a very friendly, comfortable, clear manner”
  - “basic info made easy to understand – should teach health stats”

**Conclusions: yes, worth doing, and yes, they did want to understand basics from a research point of view. And they wanted it delivered in clear, simple & friendly manner**

# Follow-up session: trial & evaluation, 2005

## ■ Session Objectives

- *Focus on how to use the key techniques of chisquare, ANOVA, and regression in detecting what is happening in a dataset, including awareness of possible problems and how to use diagnostic plots to check for problems.*
- *Techniques will be demonstrated and discussed through real datasets with interesting statistical features that are explored in a holistic way within context, emphasizing the importance of understanding the data context.*
- *This session is a follow-up to the first one, focussing on topics most commonly requested by the participants.*

**26 attended: almost unanimous approval - “very useful” “wonderful” “very good pace”; strong support for planned series of 4 symposia**

# **Symposia in Statistical Thinking for Postgraduates 2006 -**

- **Series of 4 symposia, 2.5 hours each (incl coffee break!) given twice a year.**
- **Timetable, registrations & evaluations done by Research Students Centre**
- **40-50 registrations each time; approx 65%-70% attendance**
- **Types of comments & approval ratings generally consistent over the years (approx 90% approval)**
- **Some staff in first couple of years; now mostly pg students in their first semester/year, masters as well as PhD. Across faculties**
- **Pre-symposia questionnaire retained with slight adaptations**

# Symposia in Statistical Thinking for Postgraduates

***Aim: assist pg students with statistical understanding essential to planning, carrying out and analysing data investigations using core statistical tools.***

- ***For those with some statistical background***
  - *help consolidate and improve confidence in using core statistical tools in real & sometimes challenging applications, & provide indicators of further statistical tools you may wish to learn.*
- ***For those with little statistical background***
  - *help to see what can be achieved with core statistical tools.*
- ***For others***
  - *help identify and fill in gaps in your statistical knowledge and understanding.*
- ***Series is modularised - optional questionnaire organised under the headings of the four symposia.***
- ***Real datasets in contexts that do not require discipline-specific knowledge, but sufficiently complex to allow demonstration of choice, use, interpretation and synthesis of statistical tools and thinking.***
- ***Computer output is used throughout the symposia but the symposia do not include teaching the use of a particular statistical package. (NB: Associated SPSS workshops 2007-)***

# Symposia in Statistical Thinking for Postgraduates

## Titles & key points

### 1. Planning, handling & exploring data

- Research question(s)  $\longrightarrow$  planning  $\longrightarrow$  pilot
- Identify subjects, variables, types  $\longrightarrow$  planning

### 2. Categorical data

- Estimating proportions; error of estimation; general concepts of confidence intervals
- 1 categorical variable & chisq test of given proportions
- Statistical hypothesis testing. Turning research questions into statistical questions
- Two categorical variables & testing independence
- More than two?

### 3. Continuous response, categorical predictors

- Confidence interval for mean vs. prediction/tolerance interval for individual values
- ANOVA, experimental design, interaction, residuals, assumptions, diagnostics

# Symposia in Statistical Thinking for Postgraduates

## Titles & key points

### 4. **Continuous response, quantitative predictors.**

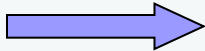
Regression (multiple)

- Models, interpreting output, residual plots & diagnostics
- Changing model
- Dependence & interaction in regression
- R-sq & cautions
- Categorical predictors?
- General linear model
- Forwards, backwards, best subsets?
- Where to go for more

Any extra discussions/points depend on group & questions but seldom arise.

- Survey questions, qualitative to quantitative, interview techniques
- Multivariate possibilities
- Repeated measures; also see SPSS workshops

# Developments & SPSS w'shops

- **More opportunity to find out areas of participants & orient the important ad hoc examples (also important in teaching MBA students)**
  
- **Aspects known to be needed from start require even more attention but benefits are great**
  - **Variables & their types, subjects**
  - **Scientific hypotheses vs. statistical hypotheses**
  - **Scientific experiments vs. statistical investigations**
  - **Essential roles of graphs in exploration & diagnostics**
  - **Dangers of one-explanatory-variable-at-a-time**
    - **Beware of t!**
    - **Beware of correlations!**
  - **Coping with discipline requirements/habits without compromising statistics**
  
- **Three SPSS hands-on workshops linked with & immediately following symposia topics 2, 3 & 4**
  - **Include repeated measures**
  - **Limited to 30 registrants  < 20 participants**

# Some pre-symposia questionnaire results

Results for 35 questionnaires across years & faculties  
(DK = don't know or no response)

- 9 stat lit questions: scores from 2 to 9, median 6
- 2D vs 3D: 34% correct, 20% DK
- Identifying continuous variables:  
9% correct, 11% DK
- Identifying categorical variables:  
20% correct, 9% DK
- Identifying subjects: 23% correct, 11% DK
- Interpreting boxplots (out of 7):  
scores from 0 to 7, median 5
  
- Incorrect coding of N/A: 60% correct, 14% DK
- Estimating conditional probabilities (2 questions):  
80% correct  
71% correct, 6% DK



# Some pre-symposia questionnaire results

- **Situation for (chisq) test of independence**
  - Identify which test needed: 23% correct, 43% DK
  - State null: 20% correct, 31% DK
  - Interpret p-value (in general terms): 63% correct, 31% DK
- **Situation for testing mean:**
  - Identify null: 26% correct, 66% DK
  - Interpret p-value (in general terms): 57% correct, 34% DK
- **Identify problem with multiple t-tests:**  
6% correct, 51% DK
- **Compare multiple t-tests with multiple predictor ANOVA:**  
3% correct, 54% DK
- **Problem of doing 1-way ANOVA on carefully designed randomised block experiment:**  
6% correct, 71% DK

# Some pre-symposia questionnaire results

- Identify mistake in interpreting confidence interval for mean as prediction interval for individual values (quality control context):  
6% correct, 66% DK
- Identify problem in judging regression model just on p-value & R-sq:  
0% correct, 69% DK
- Identify problem in using categorical predictor (>2 categories) in regression:  
23% correct, 66% DK
- Identify problem in using ordinal variable as response in regression:  
17% correct, 71% DK

# Some participant comments

- **Fantastic! I arrived being “scared of numbers” and left the session feeling empowered.**
- **It may sound crazy but I really “enjoyed” this learning experience. Thanks.**
- **...you motivated me to want to understand**
- **I’m starting to understand.**
- **Really oriented to our needs**
- **I did not understand it (regression) but I realise now what I need to know.**
- **Really good to refresh knowledge of statistics & learn how to apply to our research.**
- **Found ANOVA interesting.**
- **Clear, hit points.**
- **I really enjoyed it. It is helping me to understand stats well.**
- **I become clearer about the purpose of using statistics for my research.**

# Some participant comments

- .....a joy to be there
- ....real-life examples of the data and the interpretation of the results was simply great
- It was fantastic. .... explained this heavy subject with appropriate examples. Now I am confident to use the techniques I had learnt in analyzing my data
- It is becoming more and more interesting. I would definitely use the knowledge in the evaluation phase of my experiments.
- Excellent resources. Very useful for research students.
- I found that it was a very good and succinct summary of the approach needed for different data sets and in different types of analysis.... .... being a science student ..... I am very happy I attended all sessions as I found them to be a very helpful revision of where and how different tests should be used and has given me enormous confidence in reading published work and also considering stats while designing my experiments.

Only a few negative comments: Pre-2007, wanted SPSS workshops

Others were from extremes: “too fast” “more time to digest” “more depth please”

## Some observations

- \* Postgraduates are students as well as researchers
- \* Ug service courses in statistics should go to  $>2$  variables ASAP & avoid focus on 2-sample & simple linear regression
  - \* Ug investigations/projects should avoid design-for-procedure
  - \* Ug must include focus on essential concepts of statistical inference & data modelling (variables, estimation vs prediction, "scientific" vs statistical)



Thank you

